Using Self Organizing Maps to Find Good Comparison Universities

Cameron Cooper[1] and Robert Kilmer[2]

[1] Fort Lewis College, Freshman Mathematics Program, Durango, Colorado, United States
cooper_c@fortlewis.edu
[2] Walden University, School of Management, Minneapolis, Minnesota, United States
robert.kilmer@waldenu.edu

Abstract. Colleges and universities do not operate in a vacuum and they do not have a lock on "best practices". As a result it is important to have other schools to use for "benchmark" comparisons. At the same time schools and their students change. What might have been good "benchmarks" in the past might not be appropriate in the future. This research demonstrates the viability of Self Organizing Maps (SOMs) as a means to find comparable institutions across many variables. An example of the approach shows which schools in the Council of Public Liberal Arts Colleges might be the best "benchmarks" for Fort Lewis College.

Keywords: Kohonen self organizing maps, neural networks, benchmarking, higher education.

1 Introduction

Competitive organizations, such as institutions of higher education, continuously need to compare how well they do in relation to other organizations of similar size, shape and function across many different dimensions. These institutions could benefit from a methodology to identify organizations that are most "like" them. The example presented in this research utilizes a publically available dataset [1], self-organized maps, an unsupervised artificial neural network, to cluster/map 25 schools across 28 variables into seven distinct clusters with an easy to interpret 2-dimensional visual map of the schools and their corresponding peers. Schools within the same clusters (usually four to five other schools) could be treated as "peer" institutions for benchmarking purposes.

2 Background

Benchmarking has been used in higher education for over two decades. [2] As a means to find good comparison universities, the authors opted to utilize a Self Organizing Map. SOMs were initially crafted by Finnish professor Tuevo Kohonen in the 1960s. A comprehensive mathematical description of Kohonen's work can be found in his book *Self-Organizing Maps* [3]. Researchers have found SOMs to be

F. Schwenker and E. Trentin (Eds.): PSL 2011, LNAI 7081, pp. 148–153, 2012.

useful in medicine for visualization of gene expression [4], [5] and numerous other medical applications. In business, they have been employed for financial benchmarking [6] and strategic positioning [7]. Additionally, SOMs create means for analysis via multidimensional scaling of highly complex data sets [8], a key feature of SOMs. The primary author also applied SOMs to identify benchmark universities on the basis of student assessment of university websites [9]. The purpose of this research is twofold. First, the SOM stemming from this research offers strategic positioning information for colleges and universities. Second, via multidimensional scaling, a key feature of SOMs, this research consolidates a 28 variable input space down to an easy to interpret two-dimensional visual map.

3 Methods

The Data Profile of the member institutions of the Council of Public Liberal arts Colleges offers a nearly complete dataset [1], one important requirement to creating a reliable SOM. Thus, the data from "Section I: Admissions and Student Characteristics" of the Data Profile [1], comprised the dataset to form a 5 X 5 Self-Organizing Map (SOM). The variables for the 25 institutions are listed Table 1.

Table 1. Admissions and Student Characteristics Variables

Variables from Section I: Admissions and Student Characteristics		
1. number of applications received	10. average ACT Composite score	20. percent out-of-state undergraduates
2. percent applicants admitted	11. average SAT Critical Reading score	21. percent undergraduates age 25 or older
3. percent admitted applicants who enrolled	12. average SAT Writing score	22. percent undergraduates living on campus
4. Number (headcount) of all full-time first-time freshmen	13. average SAT Math score	23. Number of new undergraduates who were transfer students
5. percent full-time first-time female freshmen	14. Number (headcount) of all undergraduates	24. Number (headcount) of graduate students
6. percent full-time first-time male freshmen	15. percent full-time undergraduates	25. Full-time equivalent (FTE) of all students*
7. percent minority full-time first-time freshmen	16. percent minority undergraduates	26. undergraduate FTE enrollment*
8. Number (headcount) of all part-time first-time freshmen	17. percent female undergraduates	27. graduate FTE enrollment
9. percent in top 10% of high school class	18. percent male undergraduates	
	19. percent in-state undergraduates	

The authors created a two-dimensional self organizing map via a Microsoft Excel add-in, SOMinExcel [10]. The following four steps comprised the mapping process:

1) The application initially assigns random weight vectors to the neuron centers of the 25 nodes/neurons.
2) The application calculates the Euclidean distance between all of the weight vectors and a presented observation. After the best matching unit is found (i.e. the closest distance between an observation/school and the weight vector). The school is assigned to the node with an appropriate distance from the center.
3) Then, the weight vector of the winning neuron and its neighbors are modified to bring their measures closer to the observation vector. The researchers employed a Gaussian neighborhood function for adjusting the weights of the neighboring nodes. Training comprised λ=200 cycles and ultimately formed the weighted SOM. Gaussian neighborhood updates allow the weights of the neurons closest to the winning neuron to be updated to become more similar to the weights of the winning neuron. Per the Gaussian function, the magnitude of weight modification tapers off for neurons further away from the winning neuron.
4) In the end, the software finally mapped the school records onto the SOM by determining a school's relative location within the winning neurons. This process located the plot points on the 5 X 5 map.

4 Results

Seven clusters with an average cluster size of four COPLAC member institutions resulted (see Figure 1). With the goal of finding good comparison schools, a cluster size of four offers schools' administrations a reasonable number of comparison universities.

With a two-dimensional map, institutions of higher education interested in benchmarking their operations (e.g. retention and recruitment efforts) need to only locate their school's observation number (i.e. point on the map labeled by observation number from Table 3), and find schools closest to their point in terms of Euclidean distance.

5 Discussion

For example, Fort Lewis College (Observation #2 in Figure 1), the first author's home institution, might compare itself to schools #6 (Mass. College of Liberal Arts), #4 (Henderson State University), #10 (Shepard University) and #24 (The Univ. of Virginia's College at Wise), which are within the same winning neuron. By performing a radial search from the school's labeled observation point in the SOM (see Figure 1), one would find Fort Lewis College to be most similar to Observation #4, Henderson State University. This, however, does not necessarily mean Fort Lewis College would want to model and/or compare its operations after Henderson. Per a school criteria created by administrators and school-wide initiatives, one could manually create a weighted ranking of the potential comparison schools within the same neuron.

As an example, administrators at Fort Lewis College might identify the following five variables to be the most important to improve in relation to strategic initiatives:

1. % Minority Students – percent minority full-time first-time freshmen
2. % top 10% of high school class – percent in top 10% of high school class
3. % In-state – percent in-state undergraduates
4. % 25 > age – percent undergraduates age 25 or older
5. % on campus – percent undergraduates living on campus

A weighted score out of 100 could then be created in order to rank the institutions in (Table 2) reference to desired comparability. A sample weighting with % of top 10% of high school class (i.e. 35 out of the 100 points) as the foremost priority followed by increasing the % minority (i.e. 25 points) at the institution. With these weighted priorities, it would behoove Fort Lewis College to contact Henderson State University and Massachusetts College of Liberal Arts regarding admission operations.

Per the SOM, given the innate similarities between Fort Lewis College (Observation #2) and Henderson State University (Observation #4) in regards to Admissions and Student Characteristics, one could hypothesize strategic initiatives successful at one institution would also be successful for the other.

Fort Lewis College, Henderson State University, and University and Massachusetts College of Liberal Arts are all public institutions located in different geographical areas of the United States (i.e. Colorado, Arkansas, and Massachusetts). Since they are unlikely to be competing for the same applicant pool, Henderson's Office of Admission and Massachusetts College's Office of Admission might be willing to share some of its more effective "best practices" in recruiting students from the top 10% of their graduating class.

Specifically, Fort Lewis' Office of Admission might be interested in contacting their departmental counterpart at Henderson State University in bettering its recruitment of students "in top 10% of [their] high school class", variable #9. 16% of Henderson State University's students and 21% of Massachusetts' College's students are from the top 10% graduating class whereas 7% of Fort Lewis College's students are from the top 10% of their high school class. If these benchmark schools could offer Fort Lewis College, one or two in-state recruitment strategies, then the dividends to Fort Lewis could be substantial. An increased academically prepared student body will inevitably result in higher retention rates and graduation rates, both viewed as high priority strategic initiatives supported by not only higher administration but also the board of trustees.

Table 2. Weighted Averages of Institutions within Winning Neuron

Weighting	25	35	15	15	10	Weight out of 100
Institution	%Minority	% Top 10%	% In-state	% 25 > age	% on campus	Weighted Score
Fort Lewis College	26	7	69	15	36	25.15
Henderson State Univ.	40	16	85	17	42	35.1
Mass. College of Liberal Arts	12	21	76	17	68	31.205
Shepherd Univ.	14	21	55	18	29	24.805
The Univ. of Virginia's College at Wise	18	14	95	24	35	30.75

Table 3. COPLAC Member Institutions

Obs #	School Name	Obs #	School Name	Obs #	
1	The Evergreen State College	9	Ramapo College of New Jersey	18	Truman State Univ.
2	Fort Lewis College	10	Shepherd Univ.	19	Univ. of Mary Washington
3	Georgia College & State Univ.	11	Sonoma State Univ.	20	Univ. of Minn., Morris
4	Henderson State Univ.	12	Southern Oregon Univ.	21	Univ. of Montevallo
5	Keene State College	13	St. Mary's College of Maryland	22	UNC Asheville
6	Mass. College of Liberal Arts	14	SUNY College at Geneseo	23	Univ. of Science and Arts of Okla.
7	Midwestern State Univ.	15	Truman State Univ.	24	The Univ. of Virginia's College at Wise
8	New College of Florida	16	Univ. of Alberta Augustana Campus	25	Univ. of Wisconsin- Superior
		17	Univ. of IL Springfield		

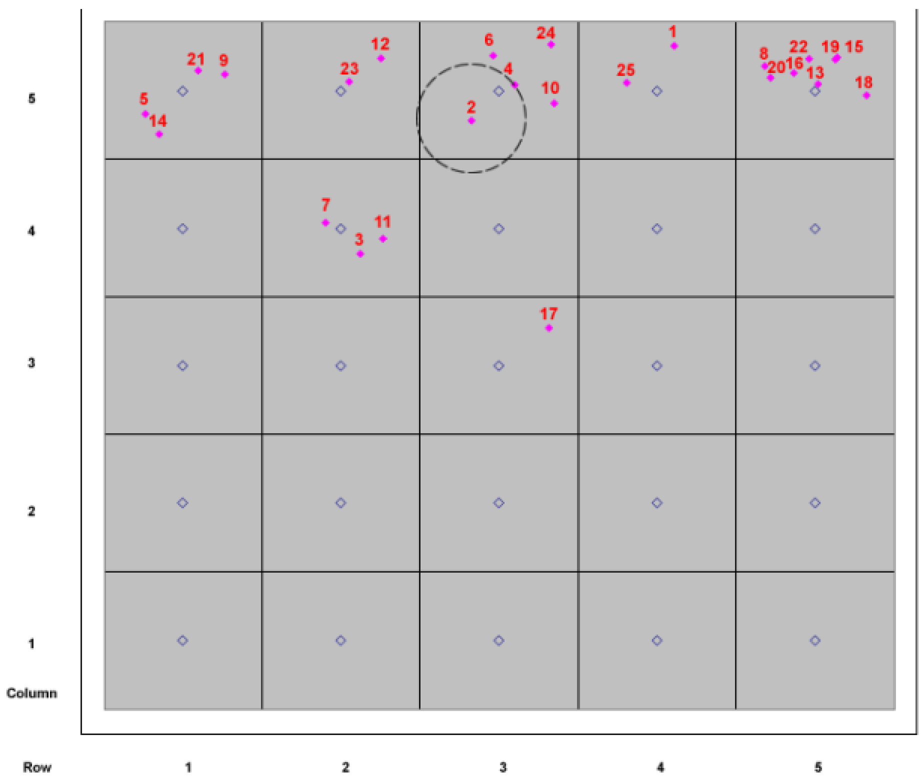

Fig. 1. Radial Search from Observation

6 Conclusions

SOMs appear to offer institutional personnel an effective, efficient, and unbiased means to discover benchmark institutions of higher education across many dimensions. Indeed the SOM in combination with a manual weighting of strategically important variables resulted in a very reasonable number of comparison schools. For future research, the results of the SOM could be used as inputs to a supervised learning approach where the examples of each cluster could be used as training and testing data.

References

[1] COPLAC, Data Profile (2009), `http://www.coplac.org/resources.html` (retrieved January 5, 2011)

[2] Benchmarking in Higher Education: Adapting Best Practices To Improve Quality. Eric Clearinghouse, Washington, DC

[3] Kohonen, T.: Self-Organizing Maps, 3rd edn. Springer, Berlin (2000)

[4] Venna, J., Kaski, S.: Comparison of Visualization Methods for an Atlas of Gene Expression Data Sets. Information Visualization 6(2) (2007)

[5] Kaski, S., Nikkilä, J., Oja, M., Venna, J., Törönen, P., Castrén, E.: Trustworthiness and metrics in visualizing similarity of gene expression. BMC Bioinformatics (2003)

[6] Visa, Eklund, T., Back, B., Vanharanta, H., Ari.: Using the Self-Organizing Map as a Visualization Tool in Financial Benchmarking. Information Visualization 2(3) (2003)

[7] Mazanec, J.: Positioning Analysis with Self-Organizing Maps – An Exploratory Study on Luxury Hotels. Cornell Hospitality Quarterly 6, 80–96 (1995)

[8] Castellani, B., Castellani, J., Spray, S.L.: Grounded Neural Networking: Modeling Complex Quantitative Data. Symbolic Interaction 26(4) (2003)

[9] Cooper, C., Burns, A.: Kohonen Self-organizing Feature Maps as a Means to Benchmark College and University Websites. Journal of Science Education and Technology 16(3), 203–211 (2007), doi:10.1007/sl0956-007-9053

[10] Saha, A.: Cited in Data Modeling for Metrology and Testing in Measurement science (2002); Pavese, F., Forbes, A.: p. 235 (2009)